VISICONTRACT: TRACING UTILITY AND IDENTITY THROUGH FEDERATED SEMANTIC CARDS

Anonymous authors
Paper under double-blind review

ABSTRACT

A client round in our federated personal-media protocol emits two representations along different trust paths: a clipped update is combined inside secure aggregation, while a typed card containing coarse time, tags, aliases, and modality confidence is read in plaintext by ingestion, storage, and card-learning components. When enabled, update differential privacy characterizes the released aggregate, not this second object. *VisiContract* turns that mismatch into an object–observer trace: it pairs channel removal with the original run, checks mechanism coverage for every realized reader and coalition, and tests attacks available from the corresponding view before comparing same-byte interfaces. In a controlled proxy built from COCO-Captions, VGGSound, and LibriSpeech-960h, blocking card consumption and its dependent objective across 1,000 non-IID clients lowers Recall@10 from 0.589 to 0.567, a paired effect of 0.022 (95% CI [0.015, 0.028]); against the strongest evaluated endpoint without a semantic uplink, FedRep, the card-enabled endpoint is higher by 0.007 [0.002, 0.012]. The useful path also retains source identity after its transport envelope is removed: alias-chain linkage assigns 0.44 of held-out cards among 100 clients, compared with 0.19 after alias removal, while mean Recall@10 changes from 0.589 to 0.585 at the matched cumulative 3.6 MB budget. Aggregate noise lowers attacks on the aggregate but cannot transform a card already visible in plaintext. VisiContract therefore identifies both the protocol object that warrants protection and the field-level intervention that directly attenuates its strongest measured identity mechanism.

1 INTRODUCTION

Follow one client upload through the server. Its update enters a secure-aggregation boundary and emerges only as part of $\widetilde{g}_r$. Its semantic card takes another edge: ingestion reads the uploader envelope and payload, strips the envelope for a downstream store, and a card learner reads the remaining time, activity, alias, and modality fields together with $\widetilde{g}_r$. The same service can therefore hold a formally covered aggregate and an uncovered plaintext representation. Calling the protocol "private" from the first edge alone assigns a property to the wrong object.

The card is not incidental telemetry in the evaluated system. Denying the card learner access to the store and removing the dependent distillation objective, while pairing client draws, update episodes, and schedules, lowers Recall@10 from 0.589 to the 0.567 FedAvg endpoint. The 0.022 difference has a paired 95% interval [0.015, 0.028]. A stronger optimizer changes the comparative endpoint but not the channel attribution: FedRep reaches 0.582 without the semantic uplink, so the card-enabled model remains higher by 0.007 [0.002, 0.012]. The update path itself also matters—FedAvg exceeds Local-only by 0.044 in mean Recall@10—making the card effect an incremental gain rather than the whole federated gain.

The plaintext representation carries a second consequence. After withholding uploader envelopes and separating attacker enrollment from query cards, alias co-occurrence links 44% of queries to the correct client in a 100-client pool; removing aliases reduces this specific attack to 19%. Rare tags and time–activity routines provide distinct linkage routes. These observations make the design question object-specific: which transmitted representation earns utility, who can read it, which mechanism covers that reader's view, and which field transformation addresses the measured attack?

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

VisiContract answers the four parts jointly. It expands a protocol into objects at component boundaries, including derived stores and coalition views; estimates conditional channel utility with paired interventions; evaluates guarantee coverage against the exact protected unit, randomization, and trust assumptions; and attaches attacks from the same view or a strict subview. A positive utility interval on an uncovered object enters a protection-priority set. Byte-matched interface and field interventions then retain the original metric units rather than collapsing utility and heterogeneous attacks into an arbitrary scalar.

This object–observer trace changes the interpretation of the experiment. The card learner's full view includes both plaintext cards and the released aggregate, so update DP remains valid for one element of that view without sanitizing the other. In the 1,000-client proxy, typed cards attain 0.589 ± 0.009 Recall@10 and preserve a +0.013 [0.012, 0.014] advantage over FedRep on unseen entity–activity–time compositions. Shuffling fields removes most of the card gain, while targeted field removals separate the utility carried by tags from the identity continuity carried by aliases. The result is a concrete protocol decision: protect the card path independently of the aggregate, and treat alias removal as a targeted defense against alias-chain linkage rather than as a protocol-wide privacy label.

2 From Update Guarantees to Multi-Channel Decisions

Personalized and heterogeneity-aware FL changes local objectives, shared layers, or aggregation under a specified interface (McMahan et al., 2017; Li et al., 2020; Karimireddy et al., 2020; Singhal et al., 2021; Arivazhagan et al., 2019; Collins et al., 2021; Li et al., 2021b; Dinh et al., 2020; Fallah et al., 2020; Li et al., 2021c;a; Oh et al., 2022; Shamsian et al., 2021; Mansour et al., 2020; Li et al., 2022; Kairouz et al., 2021). FedProto instead communicates class prototypes, making the representation itself part of the algorithm (Tan et al., 2022); FedMultimodal standardizes multimodal FL partitions and model evaluation (Feng et al., 2023). These works establish utility under their message choices. Our target is complementary: measure the conditional utility of each transmitted representation, then test whether the stated privacy mechanism binds that representation for each realized observer.

Secure aggregation, clipping, and differential privacy give precise properties for designated update releases under explicit trust and adjacency assumptions (Bonawitz et al., 2017; Abadi et al., 2016; Dwork & Roth, 2014; Mironov, 2017; Andrew et al., 2021). Empirical DP auditors test whether an implementation attains its claimed update-level bound (Nasr et al., 2023; Steinke et al., 2023). VisiContract asks a prior systems question that those audits hold fixed: which other objects are released in the same protocol, how much task utility depends on them, and whether the mechanism under audit covers their actual observer view.

Collaborative-learning attacks reveal membership and latent properties from updates (Shokri et al., 2017; Melis et al., 2019; Nasr et al., 2019; Ganju et al., 2018; Zhu et al., 2019; Yin et al., 2021); split-learning attacks demonstrate that the communicated representation determines the attack surface (Vepakomma et al., 2018; Pasquini et al., 2021). Byzantine-robust learning and poisoning attacks study integrity under malicious participants (Bagdasaryan et al., 2020; Bhagoji et al., 2019; Yin et al., 2018; Fang et al., 2020). On-device and collaborative retrieval emphasizes local relevance and reduced raw-media transfer (Ammad-ud din et al., 2019; Zhang et al., 2021; Ghosh et al., 2025). VisiContract joins the dimensions these areas hold separately: paired object utility, observer-matched mechanism coverage, and attack evidence, followed by a same-byte interface choice.

The operational distinction is direct. A protocol threat model enumerates objects and readers but need not estimate which channel drives utility. A channel ablation estimates an effect but does not determine whether a stated mechanism covers the removed object for each reader. An empirical DP audit tests a designated randomized release, whereas the failure here lies on a parallel release. VisiContract composes these measurements at one indexed object–observer pair; the resulting predicate changes a descriptive inventory into a protection priority, and its field interventions identify an actionable redesign.

3 VISICONTRACT: OBJECT–OBSERVER TRACING

3.1 COMPONENT TRACES AND OBJECT–OBSERVER PROFILES

The instantiated protocol graph has a device D, ingestion service I, semantic store S, secure aggregator A, card learner L_c, and optional admission gate G. The edge $D \to I$ carries envelope and card $(e, c^{\uparrow})$; $I \to S$ strips e but retains $c^{\uparrow}$; $S \to L_c$ supplies cards for distillation; and $A \to L_c$ supplies the released aggregate $\widetilde{g}_r$. Thus I observes $(e, c^{\uparrow})$, a non-colluding store observes $c^{\uparrow}$, and L_c observes $(c^{\uparrow}, \widetilde{g}_r, \theta)$. A coalition view is the union of its components' realized reads. This graph records every service that acquires a representation, not only the component named in its transport destination.

Let $\mathcal{V}$ denote the object channels enumerated from client-boundary sends, internal service edges, persistent stores, and broadcasts. Objects with different readers or mechanisms receive separate records; inseparable objects sharing both may be grouped. VisiContract records channel j as

$$v_j = (o_j, \mathcal{A}_j, G_j, \Delta U_j, L_j),$$

where o_j is the transmitted representation, $\mathcal{A}_j$ its protocol-realized observer set, G_j the stated guarantee with mechanism and trust assumptions, $\Delta U_j = U(\mathcal{V}) - U(\mathcal{V} \setminus \{j\})$ the paired utility contribution at the canonical configuration, and L_j a vector of attacks under declared observations. The intervention $\mathcal{V} \setminus \{j\}$ denies every consuming component access to o_j and disables computation that requires it, while holding the remaining schedule fixed. For observer a, $\kappa_j(a) = 1$ only if the guarantee's protected unit, randomization, trust assumptions, and released view all match o_j and a; a guarantee for a transformed or aggregated object does not transfer backward to its inputs. If $[\ell_j, r_j]$ is the paired confidence interval for ΔU_j, the material-gap indicator is

$$m_j(a) = \mathbf{1}[a \in \mathcal{A}_j]\, \mathbf{1}[\ell_j > 0]\, \mathbf{1}[\kappa_j(a) = 0].$$

The removal effect is a conditional marginal: random seeds, client draws, update episodes, schedules, and remaining channels are paired. Positive uncovered channels enter the decision set, where interfaces with equal semantic-byte budgets are compared on utility and every attack coordinate. Attacks may use the full observer view or a declared strict subview. Any attack using a subview remains available to the full observer, which can ignore its additional inputs; semantic-only linkage therefore applies to the card learner's richer cards-plus-aggregate view without assigning aggregate DP to the card. The output $\Pi = \{v_j : j \in \mathcal{V}\}$ preserves Recall@10, AUC, linkage accuracy, SSIM, and ASR in their native units.

For the semantic path, removing card-store consumption and its dependent objective changes Recall@10 by -0.022 (Table 3); equivalently, $\Delta U_{\text{card}} = 0.022$ with paired 95% interval [0.015, 0.028]. Ingestion I, store S, and card learner L_c observe card content, while the update mechanism has $\kappa_{\text{card}}(I) = \kappa_{\text{card}}(S) = \kappa_{\text{card}}(L_c) = 0$. The card is therefore a material uncovered object for every server-side card reader. Field removal and interface substitution then operate inside this empirically material path.

3.2 FIELD-TYPED SEMANTIC INTERVENTION SPACE

Each device stores raw media, a multimodal ANN index, and citations from retrieved moments to local identifiers and offsets. Video embeddings use sparse 1 fps keyframes plus ASR summaries. The shareable moment state is

$$c = (t_b, \mathcal{T}, \mathcal{E}, \phi, \mathcal{C}, \texttt{local_only}),$$

with coarse time bucket t_b, redacted tags $\mathcal{T}$, stable within-client aliases $\mathcal{E}$, modality statistics ϕ, local citations $\mathcal{C}$, and a local policy bit. Only $c^{\uparrow} = (t_b, \mathcal{T}, \mathcal{E}, \phi)$ may enter the plaintext card channel. Exact timestamps, free-form text, raw embeddings, locations, $\mathcal{C}$, and `local_only` remain on-device. The upload envelope $e = (u, r, n, b)$ contains a pseudonymous uploader u, round r, random card nonce n, and byte length b. Ingestion observes $(e, c^{\uparrow})$, so uploader and upload participation are direct disclosures there. The downstream store and card learner receive $c^{\uparrow}$ after e is stripped; the card learner additionally receives the released aggregate. Linkage experiments use the strict semantic-only subview, shuffle payload order within a round, and measure whether fields alone recover u. Any coalition containing ingestion inherits its direct-disclosure view instead.

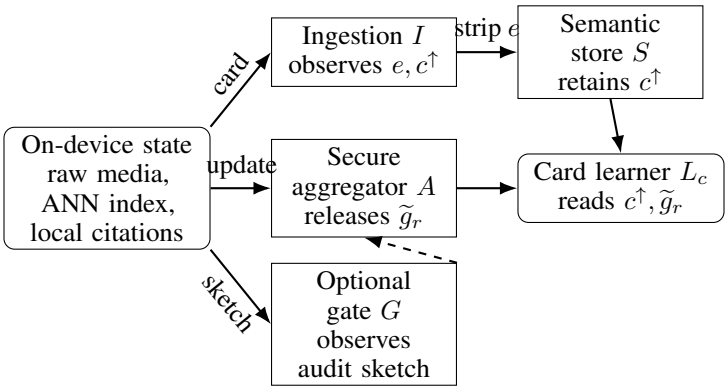

Figure 1: Object–observer trace. Ingestion sees envelope and card, the downstream store and card learner retain the plaintext card, and the card learner also receives the released aggregate. Optional sketches occupy a separate plaintext gate; raw media, ANN indices, and citations stay local.

Table 1: Moment-card schema and adjacent observation channels. Typed fields define independently removable utility and attack surfaces.

Field / channel	Server-visible?	Protocol role	Exposure tested
Transport envelope e	yes at ingestion	routing, accounting, deduplication	direct uploader and upload participation
Coarse time bucket t_b	when shared	temporal alignment in card distillation	routine and calendar-pattern inference
Redacted tag set $\mathcal{T}$	when shared	retrieval-alignment supervision	rare-activity and place linkage
Alias set $\mathcal{E}$	when shared	cross-moment entity consistency	alias-chain linkage
Modality statistics ϕ	when shared	router confidence summaries	modality and confidence fingerprints
Device-local citations $\mathcal{C}$	no	on-device evidence display	excluded from remote observation sets
`local_only` policy bit	no	blocks subsequent sharing	forward-upload control
Audit sketch (optional)	separate plaintext path	pre-aggregation participation gate	sketch exposure and adaptive evasion

Cards are sampled from the shareable local collection independently of the private gradient minibatch. Consequently, observing an upload directly reveals card contribution but not the membership target used here: whether the underlying moment participated in a local gradient episode. Typed fields make that distinction and residual semantic linkage testable. The matched interface and field-removal experiments ask whether aliases buy retrieval quality at a linkage cost, whether time buckets support routine linkage, and whether a dense alternative improves either measured axis.

3.3 Federated realization

Each selected client independently draws (i) private query–retrieval–feedback episodes for its update and (ii) shareable moments for card construction. Pretrained modality encoders produce local features; shared projection layers and the two-layer router are optimized. The private episodes remain inside the update computation. Ingestion removes e before writing $c^{\uparrow}$ to S, and L_c reads that store for $\mathcal{L}_{\text{card}}$. Separately encoded typed-field views from the same card form a contrastive positive, while views from other cards form negatives. A KL term uses the normalized uploaded modality-confidence vector ϕ as teacher and the router distribution predicted from $c^{\uparrow}$ as student. The card learner then combines this objective with the released update aggregate; it never requires a private query. Loss weights, learning rate, and step count are selected on validation clients within the shared six-trial

Algorithm 1 VisiContract-Cards round with separated communication paths

1: Card learner broadcasts shared encoder and router parameters
2: Each client samples private update episodes and an independent shareable-card pool
3: Client sends $(e, c^{\uparrow})$ to ingestion, which stores $c^{\uparrow}$ without e
4: Client clips its whole update to ℓ_2 norm C
5: Optional plaintext sketches gate participation without opening updates
6: Secure aggregation sums admitted clipped updates and injects aggregate DP noise when enabled
7: Card learner reads $\widetilde{g}_r$ and $c^{\uparrow}$, then applies update and card objectives
8: Updated shared parameters are broadcast to clients

budget. The card-path intervention blocks L_c's store reads and disables $\mathcal{L}_{\text{card}}$, while holding update episodes and schedules fixed, to estimate ΔU_{card}.

Algorithm 1 exposes four component views: ingestion sees card content and transport provenance, the downstream store sees card content without the envelope, the card learner sees that content together with the released aggregate, and the optional gate sees its plaintext sketch. Secure aggregation restricts the learner's update input; it does not restrict the learner's separate card input.

3.4 Update mechanism and observation sets

Redaction removes face-derived identifiers, proper-name and location spans, and precise timestamps before card serialization; surviving recurring entities are replaced by within-client aliases. This transformation defines card contents but is not the update-DP mechanism.

At round r, Poisson sampling with rate $q = 0.1$ selects S_r from a fixed public roster of 1,000 client slots. Neighboring federated datasets replace one slot's episodes by an empty record, so $|S_r|$ depends only on public sampling randomness. Each client's whole update g_i is clipped as $\bar{g}_i = g_i \min(1, C/\|g_i\|_2)$. Secure aggregation releases only

$$\widetilde{g}_r = \frac{1}{|S_r|}\left(\sum_{i \in S_r} \bar{g}_i + \eta_r\right), \qquad \eta_r \sim \mathcal{N}(0, \sigma^2 C^2 I), \quad C = 1.0. \tag{1}$$

Noise shares are generated inside the secure-aggregation boundary and combined before any aggregate is released; rounds below the protocol threshold abort. Thus the sum has client-level sensitivity C, and the released average has sensitivity $C/|S_r|$ conditional on the realized cohort. We compose 100 subsampled Gaussian releases with Rényi DP, $\delta = 10^{-5}$, and orders $1.25, 1.5, 2, \ldots, 512$. Noise multipliers 1.00, 1.80, and 4.30 give the reported $\epsilon \approx 8, 3, 1$ points.

The (ϵ, δ) statement binds the noisy aggregate when secure aggregation and distributed noise generation meet their threshold and non-collusion assumptions. The card learner is an authorized reader of that release, so $\kappa_{\text{update}}(L_c) = 1$ for its aggregate input and $\kappa_{\text{card}}(L_c) = 0$ for its plaintext card input. A compromised view exposing per-client $\bar{g}_i$ has $\kappa_{\text{update}} = 0$ for this aggregate mechanism and is evaluated empirically. Cards-only means $c^{\uparrow}$ plus the released shared card encoder/router, without e, update values, or sketches; it is a strict subview of L_c's cards-plus-aggregate protocol view. The canonical cards+updates diagnostic adds one compromised target-client update, the no-SecAgg intervention exposes all admitted per-client updates and participation metadata, and the DP rows expose the noisy aggregate. Cards+updates+audit belongs to the separately trained audit-enabled protocol. Appendix Table 17 fixes each attack's target and observations.

Table 2: Instantiated object–observer profile. Coverage is assigned separately to each object in a component or coalition view.

Object / channel	Realized observer view	Guarantee coverage	Utility / matched evidence
Raw media, citations	device only	remote non-upload	local retrieval reference
Envelope e	ingestion; coalitions containing ingestion	no content/privacy mechanism	direct uploader disclosure
Card $c^{\uparrow}$	ingestion; semantic store; card learner; their coalitions	no update-mechanism coverage	$+0.022\,[0.015, 0.028]$ R@10; MIA/linkage
Per-client clipped update	secure aggregator; compromised per-client view	hidden by SecAgg only outside the trusted boundary	compromised-view attacks
Released aggregate $\widetilde{g}_r$	card learner after release	aggregate RDP in DP rows	aggregate-view attacks
Audit sketch	plaintext admission gate	anomaly-screening objective	robustness bundle in Table 16
Shared parameters	card learner, clients, and model queriers	post-processing only for covered aggregate contribution	retrieval endpoint

4 Evaluation: Locating Utility and Residual Linkability

The experiments follow the object trace. First, a paired intervention estimates the incremental utility of card consumption, while no-uplink methods provide comparative endpoints. Second, attacks test direct and residual identity structure at declared component views. Third, same-byte interfaces and one-field removals locate which semantic organization carries utility and which field enables each attack. Finally, unseen event compositions, collection drift, device profiles, aggregate noise, and the audit-gated protocol test whether the resulting design decisions remain consequential beyond the canonical endpoint.

4.1 From public media to client-owned moments

Each source example becomes one client-owned moment while retaining its native media–text relation: a COCO image is paired with its captions, a VGGSound clip with its class text and audio–video content, and a LibriSpeech utterance with its transcript (Lin et al., 2014; Chen et al., 2020; Panayotov et al., 2015). Query text is held apart from the target-side representation; relevance is the paired source moment. Semantic labels drive a Dirichlet($\alpha = 0.3$) assignment to 1,000 clients, producing collections of 200–5,000 moments and non-IID activity and modality frequencies.

The common event tuple is $(\text{entity}, \text{activity}, \text{time bucket}, \text{modality})$. Activities come from source classes and content words; proper-name and location spans are removed; recurring surviving entities receive client-local aliases; a deterministic episodic schedule supplies coarse time buckets; and ϕ contains modality availability and normalized encoder confidence. Cards are created only after these transformations. The composition-shift split withholds entity–activity–time tuples from training while retaining the same clients and disjoint source items. Shadow clients used to train attacks and target clients used for evaluation never overlap.

4.2 Training, interface controls, and inference

Training runs 100 rounds with 10% participation, fp16 uplinks, one local epoch per round, batch size 32, and AdamW. The retrieval stack combines CLIP ViT-L/14, Whisper-large-v3, BGE-M3, and a two-layer 256-unit router (Radford et al., 2021; 2023; Chen et al., 2024). All utility methods share backbone initialization, client partitions, sampling schedules, query construction, and six-trial validation budgets. FedProx selects its coefficient from $\{10^{-3}, 10^{-2}, 10^{-1}\}$; SCAFFOLD maintains client and server control variates; FedRep alternates shared-representation updates with a local

projection and router head; and Ditto trains a regularized client model beside the global model. The same validation clients select learning rate and step count.

The 3.6 MB card and 11.7 $\pm$ 2.1 MB update figures are cumulative per-client uplinks over 100 rounds. Interface controls hold semantic bytes, server steps, card count, and attacker budget fixed. Shuffled cards independently permute each field across the round batch, preserving field marginals and serialization while breaking cross-field alignment. Tags-only and alias-free cards mask removed slots and pad them with ignored zeros. The dense control replaces the typed serializer with an fp16 learned projection of the same local modality features, with dimensionality chosen to match 3.6 MB. The paired no-distillation intervention supplies the conditional card-path estimate; FedRep and the other no-uplink methods supply end-to-end comparative endpoints.

Reported means and standard deviations use five matched seeds varying partitions, initialization, client sampling, and attacker training. Utility aggregates 1,000 clients per seed, giving 5,000 client-seed evaluations per method. The worst-quartile gap is the within-seed mean R@10 minus mean R@10 among that seed's bottom client quartile, followed by seed aggregation. Paired 95% intervals use $\bar{d} \pm t_{0.975,4} s_d/\sqrt{5}$ over seed-level differences.

Each membership seed uses 10,000 gradient-episode member and 10,000 non-member moments from disjoint shadow and target clients. The cards-only score uses $c^{\uparrow}$, the released card encoder/router, card-encoder confidence, card-objective loss, and field rarity; auxiliary information adds corpus-frequency and candidate-client profile features. It excludes e, update values, and audit sketches. A fixed-capacity classifier and threshold are selected on shadow validation clients and reused across matched interfaces. For linkage, each of 100 target clients contributes disjoint enrollment and query-card partitions. Candidate profiles are built only from enrollment payloads; transport envelopes are removed, and 20,000 query cards per seed are assigned among those clients. Uniform random assignment has expected top-1 accuracy $1/100 = 0.01$. Rare-tag scores use inverse public frequency, alias-chain scores use cross-card alias co-occurrence, and routine scores use time–activity sequence likelihood. Property and reconstruction attacks use the row-specific update observations stated in Table 17. The canonical point is redaction, clipping, secure aggregation, no DP noise, no audit sketch, and fp16 uplink.

5 Results: Incremental Card Utility and Semantic Exposure

5.1 Paired removal attributes a 0.022 card-path increment

Table 3: Paired controls relative to canonical VisiContract-Cards. MIA deltas use the observation view induced by each configuration; no SecAgg exposes all admitted per-client updates rather than one compromised target update.

Ablation	Δ R@10	Δ MIA AUC	Δ worst-q gap
Canonical (secure aggregation, no DP or audit, fp16)	—	—	—
Without card redaction	+0.003	+0.09	+0.008
Without secure aggregation	+0.001	+0.07	+0.004
Without distillation (FedAvg only)	-0.022	-0.01	+0.038
Half clipping threshold ($C = 0.5$)	-0.008	-0.03	+0.011
Double clipping threshold ($C = 2.0$)	+0.006	+0.05	-0.003

Blocking card-store reads and disabling the card objective recovers FedAvg at 0.567 and yields $\Delta U_{\text{card}} = 0.022$, with paired 95% CI [0.015, 0.028]. The canonical plaintext payload reaches both ingestion and the card learner, and neither secure aggregation nor Eq. 1 transforms that object. The card row therefore satisfies the material-gap predicate for both readers. Redaction is also consequential: removing it gains 0.003 Recall@10 but raises MIA AUC by 0.09. Paired intervention supplies the utility attribution; the component trace supplies the guarantee-coverage decision.

5.2 END-TO-END UTILITY UNDER SHARED SCHEDULES

Table 4: Recall@10 and cumulative per-client semantic upload over 100 rounds, across 1,000 clients and five matched seeds. Higher retrieval and lower worst-quartile gap are better.

Method	Mean R@10	Median	Q1 / Q3	Worst-q gap	Semantic MB / client
Pretrain+local fine-tune	0.571 ± 0.010	0.567	0.521 / 0.608	0.118	0
Naive FL (FedAvg)	0.567 ± 0.012	0.563	0.514 / 0.604	0.127	0
Local-only	0.523 ± 0.014	0.519	0.464 / 0.558	0.163	0
FedProx	0.575 ± 0.011	0.571	0.526 / 0.611	0.110	0
SCAFFOLD	0.579 ± 0.011	0.576	0.534 / 0.615	0.102	0
FedRep	0.582 ± 0.010	0.580	0.541 / 0.621	0.098	0
Ditto	0.580 ± 0.012	0.577	0.535 / 0.620	0.091	0
VisiContract-Cards	0.589 ± 0.009	0.586	0.543 / 0.623	0.089	3.6
Centralized training reference	0.614 ± 0.008	–	–	–	–
Split learning	0.563 ± 0.011	–	–	–	–

The results separate three gains that answer different questions. Moving from Local-only to FedAvg changes mean Recall@10 by 0.044, showing substantial utility in the federated update path. The paired removal above attributes an additional 0.022 to the card path under the canonical objective. Against the strongest listed endpoint with no semantic uplink, FedRep, VisiContract-Cards is higher by +0.007, with 95% interval [+0.002, +0.012]; against pretraining plus local fine-tuning, the difference is 0.018 [0.012, 0.024]. VisiContract-Cards also has the smallest reported worst-quartile gap. Table 5 next holds the additional 3.6 MB semantic budget fixed to identify what the card representation contributes.

5.3 MATCHED SEMANTIC INTERFACES

Table 5: Byte-matched interfaces over 5,000 client-seed evaluations. Each uploads a cumulative 3.6 MB per client; MIA uses 10,000 members and 10,000 non-members per seed.

Interface	Mean R@10	Worst-q gap	Cards-only MIA	Aux-info MIA
Shuffled card	0.569 ± 0.012	0.126	0.50 ± 0.01	0.51 ± 0.02
Tags-only	0.579 ± 0.011	0.104	0.52 ± 0.02	0.56 ± 0.03
Alias-free	0.585 ± 0.010	0.092	0.52 ± 0.02	0.58 ± 0.03
Dense learned summary	0.586 ± 0.011	0.095	0.58 ± 0.03	0.66 ± 0.03
VisiContract-Cards	0.589 ± 0.009	0.089	0.53 ± 0.02	0.62 ± 0.03

Shuffling returns utility to the FedAvg range and both MIA views to chance, locating the gain in cross-field alignment rather than transmitted volume. The dense summary reaches mean Recall@10 of 0.586 at auxiliary-information AUC 0.66, while typed cards reach 0.589 and 0.62 under the same semantic-byte budget. Alias removal changes those seed-averaged means to 0.585 and 0.58. These interface means motivate field-specific diagnosis rather than a global privacy ranking: Appendix Table 8 attributes the largest mean utility change to tags, while the direct linkage tests below show which attack mechanism aliases enable.

5.4 SEMANTIC FIELDS REMAIN LINKABLE WITHOUT TRANSPORT IDENTIFIERS

Table 6: Top-1 source-client linkage after removing transport envelope e, on 20,000 query cards per seed and a 100-client pool with disjoint enrollment payloads.

Attack / control	Top-1 linkage	Successful links
Uniform random (expected)	0.01	1,000
Rare-tag linkage	0.37 ± 0.04	37142
Popularity baseline	0.17 ± 0.03	16984
Alias-chain linkage	0.44 ± 0.05	44137
Alias-free control	0.19 ± 0.03	19052
Routine linkage	0.31 ± 0.04	30963
Shuffled-card control	0.10 ± 0.02	10121

Attack training, validation, and target clients are disjoint; target-client enrollment payloads are disjoint from query cards. Against 0.01 expected random accuracy, rare-tag linkage reaches 0.37 versus 0.17 for popularity matching, alias chains reach 0.44 versus 0.19 after alias removal, and routine linkage reaches 0.31 versus 0.10 after field shuffling. With auxiliary information, cards+updates membership AUC is 0.68 ± 0.03. Ingestion already sees uploader u; the stronger residual result is that a semantic-only observer recovers source structure after that direct identifier is withheld. Because the card learner has this entire semantic view in addition to the aggregate, each semantic-only attack is also available in its full protocol view.

The semantic effect persists under composition shift. On 37,846 unseen entity–activity–time queries per seed over the same clients, VisiContract-Cards reaches 0.551 ± 0.013, compared with 0.538 for FedRep and 0.545 for alias-free cards (Appendix Table 9).

6 DESIGN CONSEQUENCES OF THE OBJECT TRACE

The utility hierarchy separates the 0.044 mean Local-only–FedAvg gain, the paired 0.022 [0.015, 0.028] card-path increment, and the 0.007 [0.002, 0.012] advantage over the strongest evaluated no-uplink endpoint. The card is an incremental, independently material path that remains uncovered for every server-side card reader.

Controls must follow that object trace. Stripping the envelope removes direct downstream provenance but leaves three semantic linkage routes above the 0.01 random rate; alias removal changes its targeted route from 0.44 to 0.19 while other fields retain their own routes. Aggregate noise reduces update-view attacks (Appendix Table 15) without transforming cards. In the audit-gated configuration, a sketch lowers backdoor ASR from 9.6% to 3.8% but enters as a new plaintext object. VisiContract thus produces per-object protection actions instead of transferring one mechanism's label across the service.

The controlled proxy combines public multimodal collections, non-IID ownership, coarse schedules, heuristic redaction, unseen compositions over known clients, and warmed-query devices. It supports paired channel removal and disjoint attack evaluation; applying the trace elsewhere re-estimates channel effects while preserving the object–observer coverage test.

7 CONCLUSION

In this 1,000-client regime, the 3.6 MB card path contributes a paired 0.022 Recall@10, remains 0.007 [0.002, 0.012] above FedRep, and supports 0.44 alias-chain linkage after transport identifiers are removed. VisiContract connects those outcomes by tracing utility, readers, mechanism coverage, and attacks at the same object. Update protection governs the aggregate; each useful semantic representation needs protection matched to its own readers and identity mechanisms.

AI USE STATEMENT

Generative AI tools assisted manuscript language, presentation, and formatting during preparation. The authors reviewed all AI-assisted material and take full responsibility for the final manuscript.

REFERENCES

Martín Abadi, Andy Chu, Ian Goodfellow, H. Brendan McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang. Deep learning with differential privacy. In *CCS*, 2016. doi: 10.1145/2976749.2978318. URL https://doi.org/10.1145/2976749.2978318.

Muhammad Ammad-ud din, Elena Ivannikova, Suleiman A. Khan, Were Oyomno, Qiang Fu, Kuan Eeik Tan, and Adrian Flanagan. Federated collaborative filtering for privacy-preserving personalized recommendation system. *CoRR*, abs/1901.09888, 2019. URL https://arxiv.org/abs/1901.09888.

Galen Andrew, Om Thakkar, H. Brendan McMahan, and Swaroop Ramaswamy. Differentially private learning with adaptive clipping. In *NeurIPS*, 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/hash/91cff01af640a24e7f9f7a5ab407889f-Abstract.html.

Manoj Ghuhan Arivazhagan, Vinay Aggarwal, Aaditya Kumar Singh, and Sunav Choudhary. Federated learning with personalization layers. *CoRR*, abs/1912.00818, 2019. URL https://arxiv.org/abs/1912.00818.

Eugene Bagdasaryan, Andreas Veit, Yiqing Hua, Deborah Estrin, and Vitaly Shmatikov. How to backdoor federated learning. In *AISTATS*, 2020. URL https://proceedings.mlr.press/v108/bagdasaryan20a.html.

Arjun Nitin Bhagoji, Supriyo Chakraborty, Prateek Mittal, and Seraphin Calo. Analyzing federated learning through an adversarial lens. In *ICML*, 2019. URL https://proceedings.mlr.press/v97/bhagoji19a.html.

Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel, Daniel Ramage, Aaron Segal, and Karn Seth. Practical secure aggregation for privacy-preserving machine learning. In *CCS*, 2017. doi: 10.1145/3133956.3133982. URL https://doi.org/10.1145/3133956.3133982.

Honglie Chen, Weidi Xie, Andrea Vedaldi, and Andrew Zisserman. VGGSound: A large-scale audio-visual dataset. In *ICASSP*, 2020. doi: 10.1109/ICASSP40776.2020.9053174. URL https://doi.org/10.1109/ICASSP40776.2020.9053174.

Jianlyu Chen, Shitao Xiao, Peitian Zhang, Kun Luo, Defu Lian, and Zheng Liu. M3-Embedding: Multi-linguality, multi-functionality, multi-granularity text embeddings through self-knowledge distillation. In *Findings of the Association for Computational Linguistics: ACL 2024*, pp. 2318–2335, 2024. doi: 10.18653/v1/2024.findings-acl.137. URL https://aclanthology.org/2024.findings-acl.137/.

Liam Collins, Hamed Hassani, Aryan Mokhtari, and Sanjay Shakkottai. Exploiting shared representations for personalized federated learning. In *ICML*, 2021. URL https://proceedings.mlr.press/v139/collins21a.html.

Canh T. Dinh, Nguyen H. Tran, and Tuan Dung Nguyen. Personalized federated learning with moreau envelopes. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21394–21405, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f4f1f13c8289ac1b1ee0ff176b56fc60-Abstract.html.

Cynthia Dwork and Aaron Roth. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 2014. doi: 10.1561/0400000042. URL https://doi.org/10.1561/0400000042.

Alireza Fallah, Aryan Mokhtari, and Asuman E. Ozdaglar. Personalized federated learning with theoretical guarantees: A model-agnostic meta-learning approach. In *NeurIPS*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/24389bfe4fe2eba8bf9aa9203a44cdad-Abstract.html.

Minghong Fang, Xiaoyu Cao, Jinyuan Jia, and Neil Zhenqiang Gong. Local model poisoning attacks to Byzantine-Robust federated learning. In *29th USENIX Security Symposium*, pp. 1605–1622, 2020. URL https://www.usenix.org/conference/usenixsecurity20/presentation/fang.

Tiantian Feng, Digbalay Bose, Tuo Zhang, Rajat Hebbar, Anil Ramakrishna, Rahul Gupta, Mi Zhang, Salman Avestimehr, and Shrikanth Narayanan. FedMultimodal: A benchmark for multimodal federated learning. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 4035–4045, 2023. doi: 10.1145/3580305.3599825. URL https://doi.org/10.1145/3580305.3599825.

Karan Ganju, Qi Wang, Wei Yang, Carl A. Gunter, and Nikita Borisov. Property inference attacks on fully connected neural networks using permutation invariant representations. In *CCS*, 2018. doi: 10.1145/3243734.3243834. URL https://doi.org/10.1145/3243734.3243834.

Indrajeet Ghosh, Kasthuri Jayarajah, Nicholas R. Waytowich, and Nirmalya Roy. Augmenting personalized memory via practical multimodal wearable sensing in visual search and wayfinding navigation. In *UMAP*, 2025. doi: 10.1145/3699682.3728340. URL https://doi.org/10.1145/3699682.3728340.

Peter Kairouz, H. Brendan McMahan, Brendan Avent, et al. Advances and open problems in federated learning. *Foundations and Trends in Machine Learning*, 2021. doi: 10.1561/2200000083. URL https://doi.org/10.1561/2200000083.

Sai Praneeth Karimireddy, Satyen Kale, Mehryar Mohri, Sashank J. Reddi, Sebastian U. Stich, and Ananda Theertha Suresh. SCAFFOLD: Stochastic controlled averaging for federated learning. In *ICML*, 2020. URL https://proceedings.mlr.press/v119/karimireddy20a.html.

Qinbin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *CVPR*, 2021a. URL https://openaccess.thecvf.com/content/CVPR2021/html/Li_Model-Contrastive_Federated_Learning_CVPR_2021_paper.html.

Qinbin Li, Yiqun Diao, Quan Chen, and Bingsheng He. Federated learning on non-IID data silos: An experimental study. In *ICDE*, 2022. doi: 10.1109/ICDE53745.2022.00077. URL https://doi.org/10.1109/ICDE53745.2022.00077.

Tian Li, Anit Kumar Sahu, Manzil Zaheer, Maziar Sanjabi, Ameet Talwalkar, and Virginia Smith. Federated optimization in heterogeneous networks. In *MLSys*, 2020. URL https://proceedings.mlsys.org/paper_files/paper/2020/hash/1f5fe83998a09396ebe6477d9475ba0c-Abstract.html.

Tian Li, Shengyuan Hu, Ahmad Beirami, and Virginia Smith. Ditto: Fair and robust federated learning through personalization. In *ICML*, 2021b. URL https://proceedings.mlr.press/v139/li21h.html.

Xiaoxiao Li, Meirui Jiang, Xiaofei Zhang, Michael Kamp, and Qi Dou. FedBN: Federated learning on non-IID features via local batch normalization. In *ICLR*, 2021c. URL https://openreview.net/forum?id=6YEQUn0QICG.

Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *ECCV*, 2014. doi: 10.1007/978-3-319-10602-1_48. URL https://doi.org/10.1007/978-3-319-10602-1_48.

Yishay Mansour, Mehryar Mohri, Jae Ro, and Ananda Theertha Suresh. Three approaches for personalization with applications to federated learning. *CoRR*, abs/2002.10619, 2020. URL https://arxiv.org/abs/2002.10619.

H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *AISTATS*, 2017. URL https://proceedings.mlr.press/v54/mcmahan17a.html.

Luca Melis, Congzheng Song, Emiliano De Cristofaro, and Vitaly Shmatikov. Exploiting unintended feature leakage in collaborative learning. In *IEEE Symposium on Security and Privacy*, pp. 691–706, 2019. doi: 10.1109/SP.2019.00029. URL https://doi.org/10.1109/SP.2019.00029.

Ilya Mironov. Rényi differential privacy. In *CSF*, 2017. doi: 10.1109/CSF.2017.11. URL https://doi.org/10.1109/CSF.2017.11.

Milad Nasr, Reza Shokri, and Amir Houmansadr. Comprehensive privacy analysis of deep learning: Passive and active white-box inference attacks against centralized and federated learning. In *IEEE Symposium on Security and Privacy*, 2019. doi: 10.1109/SP.2019.00065. URL https://doi.org/10.1109/SP.2019.00065.

Milad Nasr, Jamie Hayes, Thomas Steinke, Borja Balle, Florian Tramèr, Matthew Jagielski, Nicholas Carlini, and Andreas Terzis. Tight auditing of differentially private machine learning. In *32nd USENIX Security Symposium*, pp. 1631–1648, 2023. URL https://www.usenix.org/conference/usenixsecurity23/presentation/nasr.

Jaehoon Oh, Sangmook Kim, and Se-Young Yun. FedBABU: Toward enhanced representation for federated image classification. In *ICLR*, 2022. URL https://openreview.net/forum?id=HuaYQfggn5u.

Vassil Panayotov, Guoguo Chen, Daniel Povey, and Sanjeev Khudanpur. LibriSpeech: An ASR corpus based on public domain audio books. In *ICASSP*, 2015. doi: 10.1109/ICASSP.2015.7178964. URL https://doi.org/10.1109/ICASSP.2015.7178964.

Dario Pasquini, Giuseppe Ateniese, and Massimo Bernaschi. Unleashing the tiger: Inference attacks on split learning. In *CCS*, pp. 2113–2129, 2021. doi: 10.1145/3460120.3485259. URL https://doi.org/10.1145/3460120.3485259.

Alec Radford, Jong Wook Kim, Chris Hallacy, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021. URL https://proceedings.mlr.press/v139/radford21a.html.

Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *ICML*, 2023. URL https://proceedings.mlr.press/v202/radford23a.html.

Aviv Shamsian, Aviv Navon, Ethan Fetaya, and Gal Chechik. Personalized federated learning using hypernetworks. In *ICML*, 2021. URL https://proceedings.mlr.press/v139/shamsian21a.html.

Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership inference attacks against machine learning models. In *IEEE Symposium on Security and Privacy*, 2017. doi: 10.1109/SP.2017.41. URL https://doi.org/10.1109/SP.2017.41.

Karan Singhal, Hakim Sidahmed, Zachary Garrett, Shanshan Wu, Keith Rush, and Sushant Prakash. Federated reconstruction: Partially local federated learning. In *Advances in Neural Information Processing Systems*, volume 34, pp. 11220–11232, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/5d44a2b0d85aa1a4dd3f218be6422c66-Abstract.html.

Thomas Steinke, Milad Nasr, and Matthew Jagielski. Privacy auditing with one (1) training run. In *Advances in Neural Information Processing Systems*, volume 36, pp. 49268–49280, 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/hash/9a6f6e0d6781d1cb8689192408946d73-Abstract.html.

Yue Tan, Guodong Long, Lu Liu, Tianyi Zhou, Qinghua Lu, Jing Jiang, and Chengqi Zhang. FedProto: Federated prototype learning across heterogeneous clients. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pp. 8432–8440, 2022. doi: 10.1609/aaai.v36i8.20819. URL https://doi.org/10.1609/aaai.v36i8.20819.

Praneeth Vepakomma, Otkrist Gupta, Tristan Swedish, and Ramesh Raskar. Split learning for health: Distributed deep learning without sharing raw patient data. *CoRR*, abs/1812.00564, 2018. URL https://arxiv.org/abs/1812.00564.

Dong Yin, Yudong Chen, Kannan Ramchandran, and Peter Bartlett. Byzantine-robust distributed learning: Towards optimal statistical rates. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 5650–5659, 2018. URL https://proceedings.mlr.press/v80/yin18a.html.

Hongxu Yin, Arun Mallya, Arash Vahdat, Jose M. Alvarez, Jan Kautz, and Pavlo Molchanov. See through gradients: Image batch recovery via gradinversion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16337–16346, 2021. doi: 10.1109/CVPR46437.2021.01607. URL https://openaccess.thecvf.com/content/CVPR2021/html/Yin_See_Through_Gradients_Image_Batch_Recovery_via_GradInversion_CVPR_2021_paper.html.

Haotian Zhang, Allan D. Jepson, Iqbal Mohomed, Konstantinos G. Derpanis, Ran Zhang, and Afsaneh Fazly. Personalized multi-modal video retrieval on mobile devices. In *ACM Multimedia*, 2021. doi: 10.1145/3474085.3481545. URL https://doi.org/10.1145/3474085.3481545.

Ligeng Zhu, Zhijian Liu, and Song Han. Deep leakage from gradients. In *NeurIPS*, 2019. URL https://proceedings.neurips.cc/paper_files/paper/2019/hash/60a6c4002cc7b29142def8871531281a-Abstract.html.

A Proxy Construction and Field Interventions

The proxy maps source-native media–text pairs into client-owned moments while keeping the event tuple and upload transformation explicit. Table 7 summarizes which structural property each component supplies; Table 8 then reports conditional one-field removals at the canonical configuration.

Table 7: Structural correspondences between the public-data evaluation and personal retrieval.

Target property	Evaluation representation	Trace role
Per-client ownership boundary	Each item belongs to one client and relevance is scored against that collection.	Channel utility is evaluated per owner.
Multimodal heterogeneity	Image, audio, and speech sources are partitioned non-IID with varying collection sizes.	Seeds vary partitions and tail behavior.
Repeated within-client retrieval	Episodic queries target recurring client-local moments.	Held-out compositions test semantic transfer.
Cross-user supervision pressure	Cards and updates supply distinct shared-training paths.	Matched removal isolates each path.
Recurring semantic attributes	Public event compositions instantiate entities, activities, and coarse time.	Field interventions test the corresponding upload.
Lifecycle behavior	Warmed-query measurements characterize the serving path.	Serving cost remains supporting evidence.

The proxy instantiates ownership, heterogeneity, repeated retrieval, and shared-supervision pressure. The composition-disjoint evaluation below changes event combinations while retaining the same client population.

Table 8: Field removal from VisiContract-Cards. Negative deltas mean lower utility or lower auxiliary-information MIA AUC.

Removed field	Δ R@10	Δ aux-MIA AUC	Decision consequence
Time bucket	-0.006	-0.03	pair with routine-linkage control
Aliases	-0.004	-0.04	remove to break alias continuity
Tags	-0.012	-0.02	preserve or protect the main utility field
Modality statistics	-0.003	-0.01	lowest mean utility change

These conditional effects keep all other fields present. The seed-averaged means separate field roles: tags carry the largest retrieval change, aliases provide cross-card identity continuity, and time buckets expose a distinct routine-linkage surface. The attack-specific controls in Table 6, rather than one aggregate privacy score, determine which field transformation addresses each mechanism.

B Composition-Disjoint Episodic Retrieval

We hold out entity–activity–time combinations from training, construct targets without item overlap, and evaluate 37,846 queries per seed over the same 1,000 clients. Across five seeds this yields 189,230 query evaluations per method.

Table 9: Retrieval on unseen event compositions. Means and standard deviations are across five seeds; lower worst-quartile gap is better.

Method / interface	Episodic R@10	Client median	Worst-q gap
Local-only	0.486 ± 0.017	0.481	0.188
FedRep	0.538 ± 0.014	0.534	0.146
Tags-only	0.535 ± 0.015	0.531	0.142
Alias-free	0.545 ± 0.014	0.542	0.128
VisiContract-Cards	0.551 ± 0.013	0.548	0.121

Using the per-seed success rates in Table 10, the paired VisiContract-Cards–FedRep difference is +0.013 with 95% interval [+0.012,+0.014]. Against alias-free cards, the difference is +0.006, interval [+0.005,+0.007].

Table 10: Per-seed episodic queries with a successful top-10 retrieval; every seed has the denominator in Table 9.

Method / interface	Seed 1	Seed 2	Seed 3	Seed 4	Seed 5
Local-only	17485	18128	18393	18809	19150
FedRep	19642	20096	20399	20740	20929
Tags-only	19491	19983	20248	20664	20853
Alias-free	19907	20361	20664	21005	21194
VisiContract-Cards	20172	20550	20891	21269	21383

Every seed preserves the ordering VisiContract-Cards $>$ alias-free $>$ FedRep in successful top-10 retrievals. The split supports generalization to unseen event compositions, while its client population remains the one used for training.

C Supporting Retrieval and Systems Results

This section reports client-local relevance, collection drift, serving cost, communication, update-noise utility, and the optional audit-gate configuration. Each result uses its own stated object and observation view; the card-path materiality conclusion remains anchored by Table 3.

C.1 Personal relevance and collection drift

The personal-query slice measures retrieval inside each client's collection, complementing the all-query endpoint. The drift sequence then follows the same trained system as collections evolve, exposing whether mean and tail-client behavior move together.

Table 11: Personal-query retrieval in the partitioned-client evaluation. Added p95 is relative to round-0 warmed-query serving.

Method	Personal R@10	Rounds to +0.01 [p25,p75]	Added p95 / worst-q gain
Local-only	0.546 ± 0.013	11 [7, 16]	$+43 \pm 12$ ms / +0.038
Naive FL	0.591 ± 0.011	32 [24, 46]	$+14 \pm 6$ ms / +0.051
VisiContract-Cards	0.607 ± 0.009	21 [15, 30]	$+17 \pm 7$ ms / +0.073

Personal-query Recall@10 is evaluated on the client-local query slice and is therefore distinct from the all-query 0.589 result in Table 4. VisiContract-Cards improves both average personal relevance and the worst-client quartile under the same partitioned ownership structure.

Table 12: Continual drift under evolving collections. Average retrieval decays gradually while the worst-client tail widens.

Month	Recall@10 mean±std	Regression rate (%) mean±std	Worst-quartile gap (pts)
0 (initial)	0.589 ± 0.009	0.0	0.089
1	0.586 ± 0.010	0.8 ± 0.3	0.097
3	0.582 ± 0.011	1.6 ± 0.5	0.112
6	0.577 ± 0.012	2.4 ± 0.7	0.128

The regression rate is the fraction of month-0 successful queries that fall outside the top ten after each collection update. Its increase and the widening worst-quartile gap show that drift reaches tail clients faster than the mean indicates.

C.2 Serving and communication paths

The serving measurements characterize warmed local retrieval, whereas the communication sweep transforms the update path. Keeping these resources separate avoids interpreting update compression as a transformation of the card fields.

Table 13: On-device warmed-query profiling. Latency excludes network I/O and assumes local embeddings and indices are already built.

Device	p95 latency (ms)	Energy / footprint
Pixel 8 Pro (phone)	397 ± 42	$2{,}140 \pm 310$ mJ; 1.8 GB
MacBook Air M3 (laptop)	218 ± 28	$1{,}260 \pm 180$ mJ; 3.4 GB
Jetson Orin NX (edge)	193 ± 24	940 ± 130 mJ; 6.2 GB

Footprints include inference assets, the run's local ANN index, and device-specific runtime dependencies. Jetson's larger deployment bundle coexists with lower latency because the footprint and execution paths measure different resources.

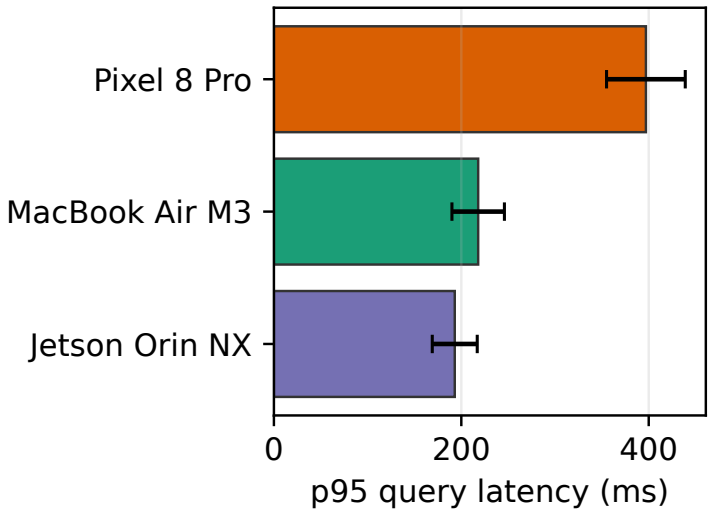

Figure 2: Per-device warmed-query p95 latency with run-to-run variation. Error bars summarize repeated query runs.

All three profiled devices answer within one second on the warmed path; the phone has the highest mean and run-to-run variation.

Table 14: Communication compression trade-offs. MIA uses the update-inclusive view; fp16 is the default operating point.

Update encoding	Uplink/client / 100 rds (MB)	Recall@10 drop	Membership AUC
fp32	22.1 ± 3.8	0.000	0.58 ± 0.02
fp16 (default)	11.7 ± 2.1	0.002	0.57 ± 0.02
8-bit quantization	7.3 ± 1.4	0.007	0.57 ± 0.02
Top-k (0.5%)	4.1 ± 0.9	0.014	0.58 ± 0.03

fp16 halves update volume relative to fp32 with a 0.002 Recall@10 drop. The near-flat update-inclusive MIA column separates communication encoding from the field transformations that alter semantic exposure.

C.3 Aggregate noise and the admission-gate path

The aggregate sweep varies the mechanism applied inside A; the audit-gated configuration instead introduces a sketch observed by G. Their juxtaposition demonstrates why two privacy or robustness mechanisms in one protocol still require separate object records.

Table 15: Utility and attack metrics under row-specific update views. Non-DP rows use their declared per-client view; DP rows expose the noisy aggregate. "n/e" means cards-only was not separately evaluated; † is updates-only because Naive FL has no card path.

Setting	R@10	Cards-only MIA	Update-view MIA	Property acc.	Recon. SSIM
Naive FL	0.567 ± 0.012	n/a	$0.69 \pm 0.03^{\dagger}$	0.74 ± 0.04	0.31 ± 0.06
Redaction + clip.	0.589 ± 0.009	0.53 ± 0.02	0.57 ± 0.02	0.61 ± 0.03	0.14 ± 0.04
DP aggregate ($\epsilon \approx 8$)	0.584 ± 0.010	n/e	0.55 ± 0.02	0.58 ± 0.03	0.11 ± 0.03
DP aggregate ($\epsilon \approx 3$)	0.578 ± 0.011	n/e	0.54 ± 0.02	0.56 ± 0.03	0.09 ± 0.03
DP aggregate ($\epsilon \approx 1$)	0.562 ± 0.013	n/e	0.52 ± 0.02	0.54 ± 0.03	0.07 ± 0.02

Within the aggregate-DP rows, attack metrics decrease monotonically as noise increases. The canonical view comparison below isolates a different question: how much one compromised per-client update adds to an attacker that already sees the cards.

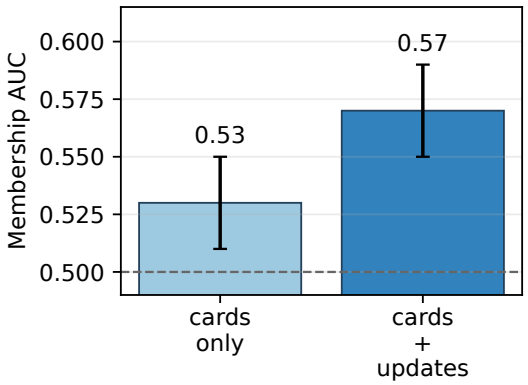

Figure 3: Same canonical model under two attacker views: semantic cards only, or cards plus one compromised target-client update. Both include the released shared card model; error bars are seed standard deviations.

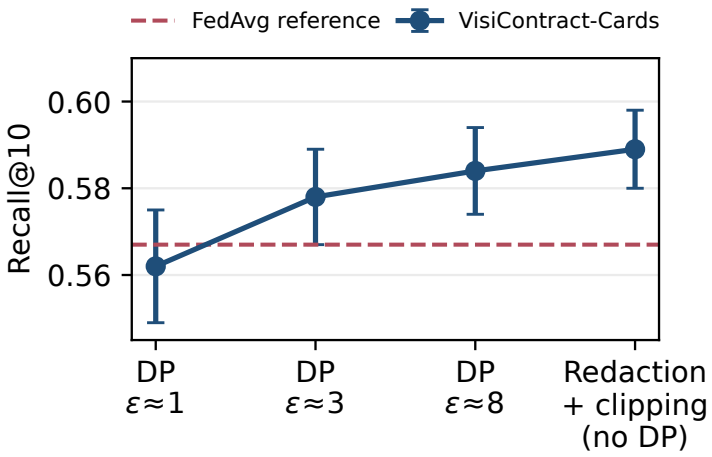

Figure 4: Update-DP operating points corresponding, left to right, to $\epsilon \approx 1, 3, 8$ and canonical no-DP redaction+clipping in Table 15. The FedAvg line is a utility reference; cards remain outside the update accountant at every point.

Table 16: Reported protocol bundles under a 10% attacker fraction and text-tag trigger, using 25 runs and 25,000 benign admission decisions.

Configuration	Backdoor ASR (%) mean±std	Recall@10 drop	Worst-q. gap / benign reject
Naive FL	24.7 ± 4.2	0.031 ± 0.008	0.138 / 0.1%
Clipping only	13.8 ± 2.9	0.016 ± 0.005	0.114 / 0.4%
Secure agg + clipping	9.6 ± 2.1	0.012 ± 0.004	0.103 / 0.6%
+ anomaly auditing	3.8 ± 1.4	0.007 ± 0.003	0.091 / 1.2%

All rows use a common finite-value and update-norm sanity gate, which accounts for the nonzero benign-rejection entries before anomaly auditing. The final comparison holds secure aggregation and clipping fixed: adding the sketch gate changes ASR from 9.6% to 3.8%, with a 1.2% benign-rejection rate. The earlier rows are protocol bundles, so their differences are not attributed to secure aggregation in isolation. The sketch remains a separate plaintext observation.

D Attack Protocol Details

The tables below make target units distinct from transport disclosure. Card linkage removes e and uses disjoint target-client enrollment payloads; membership targets private gradient-episode participation; and aggregate-DP accounting is assigned only to the released aggregate. A card learner can run every semantic-only attack by ignoring its aggregate input, while a compromised per-client update defines a different empirical view.

Table 17: Threat-model summary. Each outcome is interpreted only for its declared target and observation set.

Attack	Observation set	Reported metric
Membership inference	payload plus released card model; optionally one target update, all updates without SecAgg, the DP aggregate, or the separate audit protocol	membership-inference AUC
Property inference	per-client updates in non-DP rows; released noisy aggregate in DP rows	protected-attribute accuracy
Embedding reconstruction	per-client updates in non-DP rows; released noisy aggregate in DP rows	reconstruction SSIM
Model poisoning	text-tag trigger with 10% attacker fraction	backdoor ASR, false reject, utility drop
Rare-tag linkage	tags with e removed, plus public popularity statistics	top-1 client linkage
Alias-chain linkage	repeated aliases across payloads with e removed	top-1 client linkage
Routine linkage	time buckets and activity sequences with e removed	top-1 client linkage

The first table separates the object observed by an attacker from the property being inferred. The next table fixes the positive class, training split, and interpretation for each metric so that direct provenance, semantic linkage, and update leakage cannot be exchanged in the argument.

Table 18: Attack targets, observation sets, and evaluation units. Card attacks share one canonical model and disjoint attacker-training, validation, and target clients.

Metric / attack	Target unit and positive class	Attacker training or visibility assumption	How to interpret the reported number
Membership inference AUC	One client-owned moment; positive means it participated in a private gradient episode.	Cards-only includes $c^{\uparrow}$ and the released card model; shadow clients are disjoint. Added update views are row-specific.	Episode-member/non-member separability for the declared view.
Property inference accuracy	One client with a fixed binary collection-profile attribute.	Non-DP rows expose per-client updates; DP rows expose only the noisy aggregate. Aggregate ϵ binds only the latter.	Empirical update-view exposure of collection structure.
Reconstruction SSIM	One held-out local mini-batch designated as the reconstruction target.	Non-DP rows expose its update directly; in DP rows its contribution is contained only in the released noisy aggregate.	Comparative reconstruction fidelity within each declared view.
Backdoor ASR	One text-tag-trigger training protocol with 10% attackers.	Attackers manipulate 20 rounds; audit-enabled runs reveal a small sketch.	Trigger success paired with utility and benign false rejection.
Rare-tag linkage	One query payload assigned to one of 100 target clients.	The semantic-only attacker compares tags with disjoint enrollment profiles and public popularity after e is removed.	Residual source linkage versus popularity matching.
Alias-chain linkage	One query payload assigned through repeated aliases.	Alias co-occurrence is computed between disjoint enrollment and query payloads after e is removed.	Residual linkage versus the alias-free interface.
Routine linkage	One query card, scored with its held-out time–activity context, assigned to a target client.	Time–activity order is compared with disjoint enrollment profiles after e is removed.	Residual linkage versus shuffled cards.

Keeping target units explicit prevents direct upload disclosure, residual semantic linkage, episode membership, and update reconstruction from being merged into one leakage claim.

E Supporting Observation Set and Asset Terms

The asset table records the public datasets and pretrained encoders used to construct the proxy evaluation. The following figure then separates observations from protocol changes: the first two bars are views of one canonical model, while the audit-enabled bar comes from a separately trained protocol.

Table 19: Licenses and governing terms for reused datasets and pretrained models.

Asset	Role	License / governing terms
COCO-Captions	image-caption proxy data	COCO Terms of Use; underlying images retain their original Flickr-source licenses, so use follows the per-image source license recorded by COCO rather than a single blanket image license
VGGSound	audio-video proxy data	Creative Commons Attribution 4.0 (CC BY 4.0); original video copyright remains with the source owners
LibriSpeech-960h	speech proxy data	Creative Commons Attribution 4.0 (CC BY 4.0), derived from LibriVox public-domain audiobooks
CLIP ViT-L/14	vision encoder	MIT License, used as a pretrained research backbone without redistribution changes in this submission
Whisper-large-v3	audio encoder / ASR backbone	Apache License 2.0, used as a pretrained research backbone under the published model-card terms
BGE-M3	text encoder	MIT License, used as a pretrained research backbone under the published model-card terms

The card model and proxy assets are common to the canonical bars. The hatched bar belongs to the audit-gated extension, whose sketch adds a distinct observer and therefore a distinct object record.

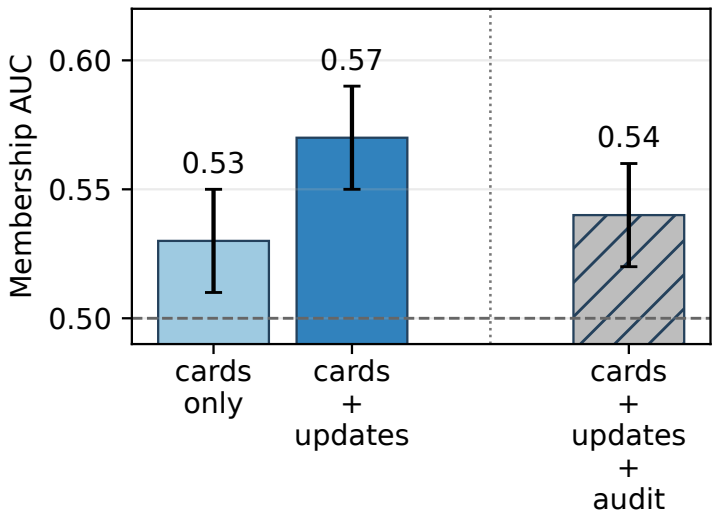

Figure 5: Protocol-grouped MIA diagnostic. The left group is the canonical no-audit model; the right hatched bar is a separately trained audit-enabled model exposing the sketch channel.

The separated grouping prevents the audit-enabled result from being interpreted as another observation set on the canonical model. Dataset terms and model licenses are listed separately because public proxy media and pretrained encoders enter under distinct governing terms.